



# Structures and functions of complex evaluation systems: comparison of six Central and Eastern European countries

International Review of Administrative

Sciences

1–19

© The Author(s) 2021



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/00208523211026964

[journals.sagepub.com/home/ras](https://journals.sagepub.com/home/ras)**Tomasz Kupiec** 

University of Warsaw, Poland

**Dominika Wojtowicz** 

Kozminski University, Poland

**Karol Olejniczak** 

SWPS University of Social Sciences and Humanities, Poland

## Abstract

Evaluation practice is vital for the accountability and learning of administrations implementing complex policies. This article explores the relationships between the structures of the evaluation systems and their functions. The findings are based on a comparative analysis of six national systems executing evaluation of the European Union Cohesion Policy. The study identifies three types of evaluation system structure: centralized with a single evaluation unit, decentralized with a coordinating body and decentralized without a coordinating body. These systems differ in terms of the thematic focus of evaluations and the targeted users. Decentralized systems focus on internal users of knowledge and produce mostly operational studies; their primary function is inward-oriented learning about smooth programme implementation. Centralized systems fulfil a more strategic function, recognizing the external audience and external accountability for effects.

---

## Corresponding author:

Tomasz Kupiec, University of Warsaw – Centre for European Regional and Local Studies, Krakowskie Przedmieście 30, 00-927 Warsaw, Poland.

Email: [tomasz.kupiec@uw.edu.pl](mailto:tomasz.kupiec@uw.edu.pl)

## **Points for practitioners**

Practitioners who design multi-organizational evaluation systems should bear in mind that their structure and functions are interrelated. If both accountability and learning are desired, the evaluation system needs at least a minimum degree of decentralization on the one hand and the presence of an active and independent coordination body on the other.

## **Keywords**

accountability, cohesion policy, evaluation, evaluation systems, evaluation functions, learning, structures

## **Introduction**

Evaluation, defined as a systematic inquiry of the worth and merit of public interventions (Fournier, 2005; Patton, 2004), has been considered a crucial element of the policy cycle and public administration practice for at least several decades (Chelimsky and Shadish, 1997). Its roots can be traced to the United States' War on Poverty and Great Society initiatives in the 1960s (Mark et al., 2006; Rossi et al., 2004). The subsequent waves of public management reforms, including performance-based management promoted by the Organisation for Economic Co-operation and Development and the World Bank (Kusek and Rist, 2004; OECD, 1998), institutionalized evaluation within governments and made it a ubiquitous practice across public administrations (Furubo et al., 2002; Stockmann et al., 2020).

In Europe, the European Union (EU) has been the major promoter of evaluation practice in public administrations at various levels. While some EU Member States have developed their own evaluation culture, during what is often referred to as the first and second evaluation wave (Derlien, 1990), many others, including all those joining the EU in the twenty-first century, adopted this practice mainly in response to the requirements of EU regulations.

The ultimate and general goal of evaluation is 'social betterment' (Henry and Mark, 2003). In the practice of specific organizations, evaluation translates into various functions (also called purposes), often driven by conflicting logic, and therefore leading to potential tensions among the actors in policy systems (Donaldson et al., 2010). Thus, it is important for both the theory and practice of evaluation in public administration to recognize the functions underlying specific evaluation efforts and to identify their determinants.

The vast majority of evaluation use literature focuses on a single evaluation study as the unit of analysis (Højlund, 2014; Raimondo, 2018) and misses the wider institutional context. Only a handful of authors examine evaluation systems in terms of multi-organizational networks (Olejniczak, 2013; Stockmann et al., 2020), but they do not explore, in a comparative manner, the relation between structural arrangements and system functions. Some studies suggest that there might be a relationship between

functions and the design of evaluation systems. However, these articles take the perspective of an individual organization and their internal evaluation arrangements. A group of authors postulate that a broader perspective of evaluation systems is needed, and more attention should be devoted to studying the arrangement of evaluation systems as multi-organizational networks (e.g. Kupiec, 2020; Leeuw and Furubo, 2008; Liverani and Lundgren, 2007).

This limited recognition of multi-organizational systems and their structures in evaluation literature is in sharp contrast with the wider body of work on administration and management. The effects of institutional structures on political and administrative outcomes lie at the heart of the field of public administration and policy (Balla et al., 2015). Management literature also acknowledges the association between functions (strategy) and structure (Dyas and Thanheiser, 1976; Mintzberg, 1992).

Therefore, there is a gap to be addressed by exploring the relations between structures of multi-organizational evaluation systems and the functions they fulfil. This study defines the evaluation system as a set of organizations involved in acquiring or using evaluative knowledge, while the structure is the relationships between them. Our hypothesis states that the structure of evaluation system correlates with its functions. We do not intend to establish the direction of this relationship, but we claim that certain structures may support one function and impede another, thus indicating the dominant orientation of a specific evaluation system.

To verify this hypothesis, we analyse the evaluation practices of six EU Member States that are substantial beneficiaries of cohesion policy (CP): Bulgaria, Czechia, Hungary, Poland, Romania and Slovakia. The evaluation practice in CP is considered to be the most developed among EU policies (Fratesi and Wishlade, 2017). This is mostly due to the significant size of the CP budget, the spectrum of activities that it covers and the long record of making evaluation an integral part of programme management. As longitudinal analysis shows, these extensive evaluation practices prescribed by the European Commission (EC) result in a spectrum of evaluation functions, with some trade-offs among functions (Batterbury, 2006).

Although the EC introduced the same requirements for monitoring and evaluation in all Member States (regulation no. 1303/2013), it left open the issue of structural and procedural arrangements. As a result, national administrations developed country-specific evaluation systems with distinctive structures and operating procedures. This situation creates a unique opportunity for a comparative study.

Even though it has not been explicitly stated in the regulations, CP evaluation arrangements seem to be based on the equifinality principle, which states that the same evaluation functions can be performed by various, national-specific systems. This article challenges this implicit assumption by arguing that national evaluation systems of varying structures perform different functions, and therefore common CP evaluation goals are not achieved in all countries to the same degree.

The article commences with a review of evaluation functions in public administration, summarizing them with a two-dimensional, four-element model of evaluation functions. This synthesis provides a frame for later empirical study. A discussion on the boundaries of evaluation systems is also provided in this section, concluding with the concept of a complex multi-organizational evaluation system. We then move on to explain the

research design and methods, including the rationale behind the selection of national cases, the approach to operationalizing and measuring key variables and sources of data. The next part is devoted to a comparative analysis of the variations in structures, activities and functions of six national evaluation systems. The article concludes with implications for future research and evaluation practice.

The article aims to refocus the evaluation debate onto a new unit of analysis – evaluation systems understood as multi-organizational arrangements – and provide an initial typology of system structures. We hope that this will pave the way for a more systematic inquiry into the relationships between evaluation functions and structures as part of public administration and policy practice.

## Theoretical framework

### *Evaluations functions*

Evaluation is an essential element of a policy process and public administration practice. Although policy cycle models differ from author to author, they always include an evaluation stage. Evaluation contributes to the improvement of public policies and ultimately to social betterment (Henry and Mark, 2003). Its importance for policy making, public administration and, more generally, the public sector is explained in many various, complementing ways. For Ostrom (2007) evaluation is necessary for the efficient and just management of common pool resources, and Picciotto (2016) argues that from the perspective of agency problem, evaluation reduces information asymmetry. Reducing red tape is also among the highlighted benefits (Oh and Lee, 2020). The use, non-use, advantages and limitations of evaluation are discussed in the context of bureaucracies and political authorities (Kudo, 2003), local governments (Favoreu et al., 2015) and national parliaments (Speer et al., 2015).

A rich body of literature describes a wide range of potential evaluation functions also called purposes (e.g. Batterbury, 2006; Chelimsky and Shadish, 1997; Hanberger, 2011; Mark et al., 1999; Widmer and Neuenschwander, 2004; Zwaan et al., 2016). The following observations emerge from that literature.

First, evaluation functions may be grouped into general categories (Kupiec, 2020):

- providing an assessment of performance for the external audience – *accountability, oversight, compliance*;
- inducing a change in behaviour, adjustment of a policy/strategy/programme/practice – *learning, improvement (of performance, planning), knowledge creation, building capacity*;
- supporting or criticizing a policy, practice or decision based on other (than evaluation) considerations, building appearances of a learning organization – *legitimizing, substantiating, justifying, sanctioning, political ammunition*;
- involving stakeholders in the evaluation process, ceding power to those normally excluded from the decision-making process, fostering public debate – *engaging, empowering, developing a sense of ownership*.

Second, two purposes – accountability and learning – are often considered the main ones (e.g. Van Der Meer and Edelenbos, 2006). Rather than being complementary, these two aspects stand, to some extent, in opposition to each other. While learning is oriented inward (to an institution, which implements an intervention subject to evaluation), accountability is performed for other stakeholders, i.e. supervising bodies, media, society. Each function requires a different attitude from evaluators, and as Raimondo (2018) argues, focusing on accountability strengthens norms which then impede learning.

Third, there are some ambiguities in terminology. The terms purposes and functions are often used interchangeably (e.g. Mark et al., 1999) without specifying their focus-specific content (i.e. accountability for what actions, learning about what issues, etc.).

Therefore, for our analytical framework, we propose referring to accountability and learning as ‘evaluation purposes’, not functions. Furthermore, the second dimension is introduced – the focus of evaluation activities. It is based on a classic dichotomy from management literature – the strategic vs. operational focus. Combining the purposes with focus allows four possibilities to be identified. We call them ‘evaluation functions’ (Figure 1).

The framework is aligned with the accountability literature distinguishing between accountability for inputs and process vs. outputs (Dubnick and Frederickson, 2011) and the organizational learning literature – operational improvements in ‘doing things right’ or strategic reflection on ‘doing right things’ (Argyris and Schon, 1995). The framework also corresponds to the goals of CP evaluation specified in EU regulations: (1) supporting the management of the programmes (operational and, to a lesser extent, strategic learning); and (2) assessing their effects (strategic and, to a lesser extent, operational accountability). The framework provides the basis for defining variables in this study.

### *Evaluation systems – boundaries and determinants of functions*

While there is a rich body of literature on determinants of evaluation use, and specific types of use, there has been little discussion about the factors conditioning evaluation functions, especially when restricted to evaluation systems alone. Available sources suggest one potential factor – the design/structure of the evaluation system. Hanberger (2011), who studied international organizations, suggests that the evaluation functions depend on the system design. Widmer and Neuenschwander (2004) argue that the organizational design of a system should follow the purpose/function of evaluation but admit that this relationship is often two-way in the reality of Swiss public agencies.

These empirical observations correspond to some extent with findings from the management field on the association between strategy and structure. The relationship between these two aspects has been the subject of numerous studies since the 1970s, with the majority suggesting the impact of strategy on structure (e.g. Dyas and Thanheiser, 1976; Mintzberg, 1992). There is, however, also a substantial body of literature supporting the opposite claim, e.g. Hall and Saias (1980) suggest that structure impacts our perception and decisions, while Harris and Ruefli (2000) reveal that structure influences organizational performance.

		Purpose	
		Accountability	Learning
Focus	Strategic	<i>Accountability for strategy - results achieved</i>	<i>Learning about strategy - what works and why</i>
	Operational	<i>Accountability for operations - sound implementation</i>	<i>Learning about operations - improving implementation</i>

**Figure 1.** Four potential functions of the evaluation system.  
Source: own elaboration based on Olejniczak et al. (2017).

Building on the findings discussed above, this article explores the relationships between the structures and the functions of national CP evaluation systems. What distinguishes this study from previous research is the different definition of an evaluation system. As Williams and Imam (2007) suggest, thinking in terms of evaluation systems requires defining boundaries – deciding what lies within and outside of them. Hanberger (2011) and Widmer and Neuenschwander (2004) studied single organizations; therefore, their evaluation systems are embedded within the boundaries of a single organization. However, this is not the only possible arrangement of or perspective on an evaluation system.

The acquisition of evaluative knowledge is often performed by distinct bodies named evaluation units. They identify knowledge needs, conduct studies and then feed knowledge to the intended users (Olejniczak et al., 2016). User-units and evaluation units might be from the same organization but when they are not we are dealing with a multi-organizational evaluation system.

In complex policy settings, evaluation systems also become complex, and national CP evaluation systems are model examples. They may consist of a single national evaluation unit, but they may also contain many interrelated organizations acquiring and using evaluative knowledge. Not only one evaluation-acquiring organization may serve several user-organizations, but it is also possible that several organizations – each of them acquiring and using evaluation – cooperate in the process or are coordinated by one central body. Organizations comprising the evaluation system may operate at different levels, e.g. national or programme levels. Relations between organizations may resemble those characteristics of network structures or hierarchies, especially when several levels are involved. An organization from a higher level often coordinates, facilitates and creates a regulatory framework for the subordinate. Lower-level organizations often feed knowledge to their superiors.

This variety of possible relations, coordination, and cooperation modes draws our attention from the perspective of evaluation functions. Therefore, the following hypothesis can be put forward: *the structure of the evaluation system correlates with its functions*. The evaluation system here is defined as all organizations involved in acquiring or using evaluative knowledge in the CP context, while the structures are the relationships between those organizations.

Research design and methods

The research is based on a cross-national comparative design, focusing on the identification and explanation of potential diversities across CP national evaluation systems. This approach follows Lijphart (1971), who views the comparative method as one of the legitimate scientific methods of establishing general empirical propositions, and Sartori (1991), who sees comparison as an acceptable method for controlling variables.

Country selection

The study covers CP evaluation systems in six out of 12 Central and Eastern European Countries (CEECs) that joined the EU between 2004 and 2007: Bulgaria, Czechia, Hungary, Poland, Romania and Slovakia. The analysis level is a country, and the unit of analysis is a national CP evaluation system. Poland is the only country in the studied sample with a CP implementation system consisting of national and regional level arrangements. The regional level of Polish evaluation units was excluded from the analysis to ensure comparability between cases. Case selection was based on ‘the most similar system strategy’ (Sartori, 1991). Thus, countries with similar critical variables related to the administration system were selected, except for the phenomenon that was to be investigated, i.e. CP evaluation system structure (Figure 2).

All of the studied countries are characterized by a comparable administrative culture derived from the communist legacy, as they were part of the Soviet Block (Meyer-Sahling, 2009). After 1990, the course of the socio-political transformation process and administration development trajectory were also alike in CEECs (Ágh,

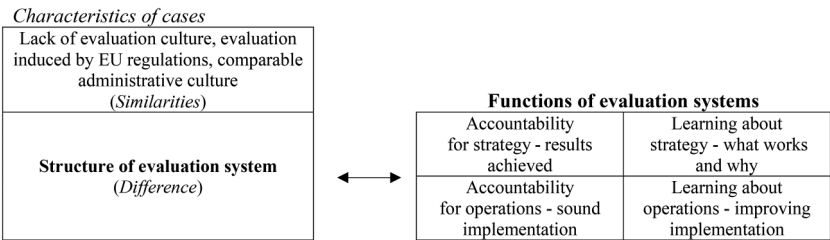


Figure 2. Research scheme.  
Source: own elaboration.

2016). Recent research proves that several common features of the communist-type administrations have persisted to the present day (e.g. Dimitrov et al., 2006) and remain reflected in the current characteristics and performance of public administration (Thijs et al., 2017). This methodological choice to pair and compare CEECs with each other is frequent in public administration studies (e.g. Ágh, 2002).

From the evaluation perspective, resemblances between the selected countries are even more evident. None of them had developed an evaluation culture before joining the EU. Hence, the evaluation practice in the CP established after 2004 was solely due to external pressure from EU regulations. Additionally, all of the studied countries are among the main beneficiaries of CP funds in per capita terms, which places specific pressure on evaluation systems.

### *Operational variables and data collection*

Table 1 summarizes how the variables of structure and dominant function of the evaluation system were operationalized and assigned measures. Data were collected at two levels: (a) national, with the focus on the structure of the system, regulations and coordination activities; and (b) organization, with the focus on potential variations in activities and the orientation of evaluation units in different national systems resulting from choices at the national level. At the organization level, the data were aggregated to provide average/dominant characteristics for the national system.

The analysed data came from the following sources:

1. A survey among the heads of all CP evaluation units in the studied countries (see Table 2). The questionnaire covered the units' organization and activities performed and the dominant target audience of executed evaluation reports.
2. In-depth interviews with representatives of coordinating bodies for CP evaluation system, leading evaluation units, local evaluation experts in each country (see Table 2). The interviews covered the activity and arrangements of the evaluation system at the national level.

**Table 1.** Operational variables and measures.

	Operational variable	Measure
Structure of evaluation system	Type of structure	Number of organizations conducting evaluation
		Degree of activity of coordination body
Dominant function of evaluation system	Type of knowledge provided by evaluation studies	Proportion of strategic/operational evaluation studies
	Dominant target audience	Share of impact-oriented evaluation studies in total
		Relative focus on external/internal institutions

**Table 2.** Data sources for analysis: survey and in-depth interviews.

Country	Survey: number of evaluation units surveyed (total number of evaluation units in the country)	In-depth interviews: number of persons interviewed
Bulgaria	5 (7)	5
Czechia	11 (12)	3
Hungary	1 (1)	4
Poland	20 (20)	3
Romania	3 (3)	6
Slovakia	13 (15)	3
Total	53 (58)	24

- 3. Desk research of English-language sources, e.g. national guidelines and evaluation plans, providing general knowledge of evaluation practices in the selected countries.
- 4. The evaluation database managed by the EC – the number of impact evaluations conducted in each country.
- 5. Six in-depth interviews with DG Regio and DG Employment representatives providing data triangulation and supplementing the national-based sources.

Descriptive statistics measures were employed to analyse sources (1) and (4). Sources (2), (3) and (5) were analysed according to the qualitative approach methodology. First, the main themes were identified with the use of coding. Patterns and relationships were then specified, which enabled the content and narrative analysis of the gathered material.

**Results**

*Description of the context – CP implementation system*

The evaluation systems that are the subject of this analysis are an inherent part of the CP implementation system. This policy is performed based on the shared management concept, which may be characterized as coordinated action taken by institutions at different levels, aimed at setting policy goals and taking concrete actions to achieve them (Dąbrowski et al., 2014). It means that responsibility and duties are shared between EU and Member States’ administrations.

The detailed scope of CP intervention and implementation arrangements in each country are negotiated between the EU and Member States and formalized in the so-called Partnership Agreement. The policy is implemented through operational programmes (OPs) covering specific areas of support. The number of OPs differs, but at least some of them are implemented in every studied country. Operational programmes are formulated and implemented by the managing authorities (MA) – usually ministries. Sometimes one institution acts as the MA for more than one OP. The MA may also delegate some of its competences to implementation bodies – other ministries or government agencies.

### *Structures and activities of evaluation systems*

Although developed under the same EU rules, CP national evaluation systems in the studied countries differ in terms of structure. The main difference is in the number of organizations that commission evaluation studies. While this activity is spread across a number of organizations in some countries – we call these systems decentralized, in the rest, evaluations are commissioned by just one organization and then fed to others – we call these systems centralized. Within the group of decentralized systems, there is another variation involving a coordination body, which exists and engages in coordinating activities in just some systems, making their structure less loose and activities more coordinated.

The most common decentralized system design can be found in Czechia, Slovakia and Bulgaria. Several evaluation units operate in each of these systems. They are located in the ministries, which act as MAs for OPs. In these systems, each evaluation unit is responsible for evaluating a single OP, implemented by the ministry that it is part of.

Poland is an example of even deeper decentralization. Evaluation units operate at the level of ministries – MAs and implementation bodies – usually government agencies responsible for implementing selected priorities under certain OP.

At the other extreme lies Hungary – a centralized evaluation system with only one evaluation unit. It operates in the Prime Minister's Office and has exclusive competences regarding the evaluation of seven OPs. The MAs for these programmes are located in different ministries. Hence, the Prime Minister's Office is the only organization commissioning evaluation, which is then fed to organizations using evaluation – ministries.

The final case, Romania, constitutes another example of a centralized structure. There is a single unit responsible for evaluation there, operating within the Ministry of Regional Development, Public Administration and European Funds. However, this unit is divided into three sub-units – two of them responsible for evaluating a single OP and the third dealing with five programmes. Implementation of all OPs is the responsibility of units within the same ministry.

The number of evaluation units corresponds to the level of decentralization of the system. A single unit in Hungary and one subdivided into three in Romania contrasts with seven units in Bulgaria, 12 in Czechia, 15 in Slovakia and 18 in Poland.

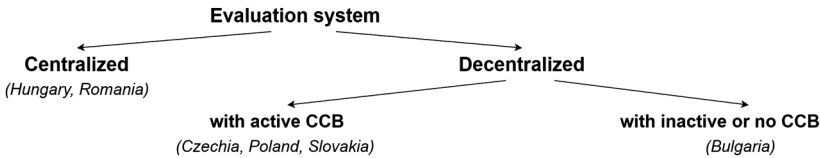
In decentralized systems, the network of organizations with evaluation units is linked by what can be termed 'the central coordination body' (CCB). The scope of activities of the CCB and the extent to which it regulates the operations in the system differ.

Active CCBs operate in Poland, Czechia and Slovakia. They establish working groups and organize meetings, trainings and postgraduate courses (Poland); they provide knowledge-sharing activities and evaluation conferences, as well as assisting in public procurement procedures. They also support the process of designing evaluation plans. To coordinate and standardize the formulation of evaluation plans, the quality assessment of completed studies and recommendation follow-up, the Polish CCB issues formal regulation, while the Czech and Slovak CCBs confine themselves to soft guidance and support. All three – Polish, Czech and Slovak CCBs – also act as evaluation units focusing on horizontal subjects and CP effectiveness at the Partnership Agreement level.

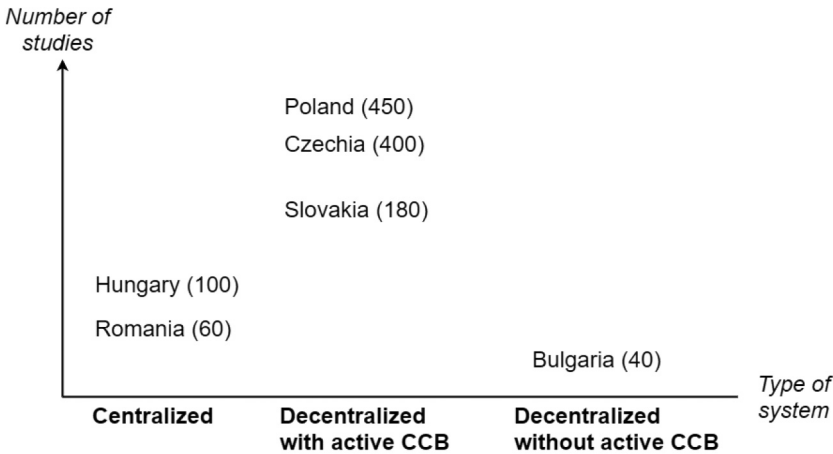
In striking contrast to the three CCBs described above is the Bulgarian CCB. Although officially present in the system, it does not perform any coordinating activities, limiting its job to merely forwarding communications from the EC and ensuring conformity to EC requirements. Bulgarian evaluation units are left to themselves when planning and conducting an evaluation or disseminating results. The CCB provides no support or training.

Based on the differences in the number of organizations with evaluation units directly acquiring evaluation knowledge, and the activity of CCB, we propose a typology of evaluation system structures consisting of three general types: centralized, decentralized with active CCB and decentralized with inactive or no CCB (Figure 3).

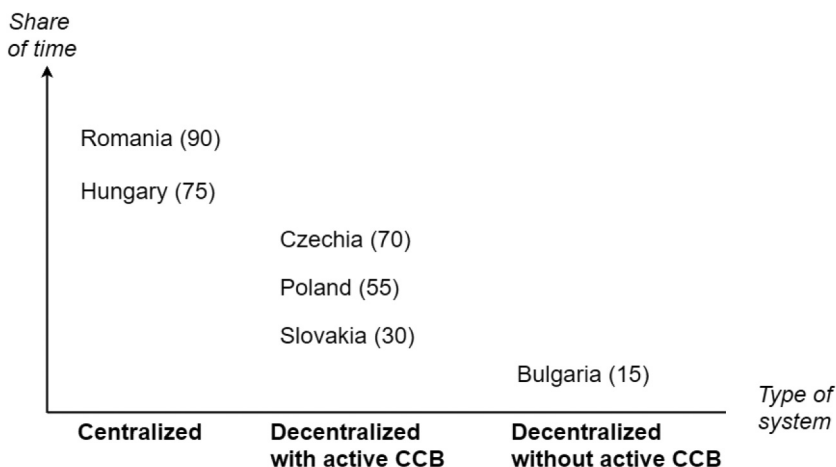
The analysis identified several differences in the activities and work organization of evaluation systems relating to their structure. The first is in the number of conducted studies. During the programming period, 2007–2013, around 100 studies were completed in a centralized system in Hungary and 60 in Romania. At the same time, 180 studies were accomplished in Slovakia, over 400 in Czechia and 450 in Poland. Bulgaria is the exception among the decentralized system with only 40 studies. Even allowing for that, the dominance of decentralized systems in terms of the number of completed studies is apparent (Figure 4).



**Figure 3.** Typology of structures of cohesion policy (CP) national evaluation systems. Source: own elaboration.



**Figure 4.** Number of completed evaluation studies. Source: interviews.



**Figure 5.** Average share of working time spent on evaluation by the staff of evaluation units (%).  
Source: survey.

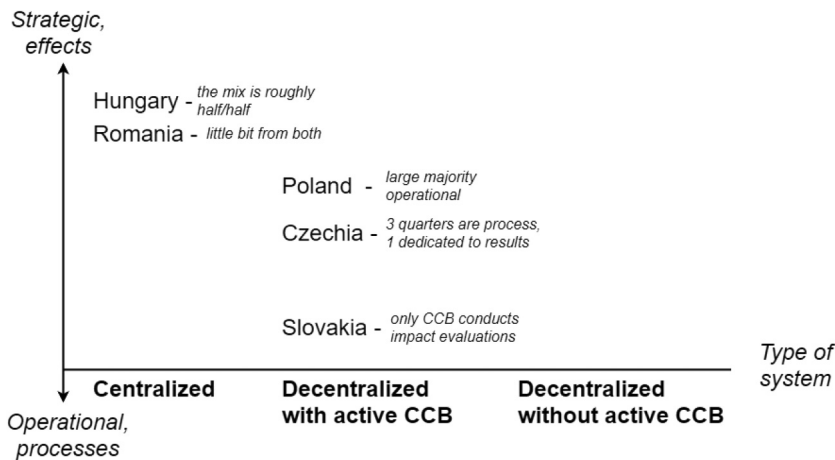
Evaluation systems also differ in terms of the average share of working time spent on tasks related to evaluation by the staff of evaluation units. In all of the studied countries, evaluation units also have other responsibilities. These include analytical activities and management activities, such as developing programmes or information and communication tasks. The proportions, however, vary greatly between countries. While other tasks amount to only 10% of working time in Romania and 25% in Hungary, they consume around 85% of working time in Bulgaria and 70% in Slovakia. Hence, in the centralized systems units usually specialize in evaluation and deal almost entirely with tasks related to evaluation studies. In decentralized systems units usually have a more diversified profile, and evaluation is just one task among others (Figure 5).

### *Functions of evaluation systems*

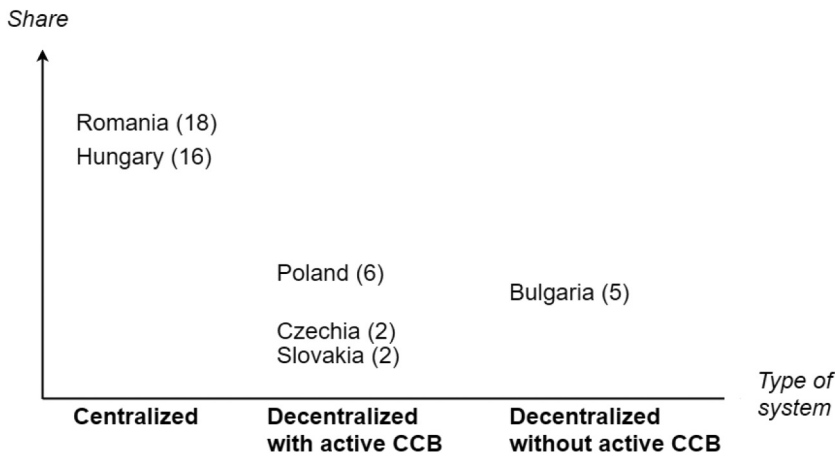
This section discusses elements of the practice of CP evaluation systems, which indicate different functions of the systems. Findings are presented for the three proposed types of structure (compare Figure 3), allowing observation of the potential correlation between the system's structure and functions.

Local experts and representatives of CCBs were asked about the proportion of strategic (focusing on the effects of intervention) and operational evaluation studies (focusing on the process of implementation) generated in their countries. According to the interviewees from Hungary and Romania, their systems provide a balanced mix of strategic and operational knowledge. In the case of all decentralized systems, operational focus dominates (Figure 6).

The observations from interviewees are confirmed by the data from the EU library of CP evaluations. Studies collected there are classified as focused on impact, monitoring/progress or implementation/process. As presented in Figure 7, impact-oriented studies



**Figure 6.** Focus of evaluation studies – perception of local experts. Bulgaria – no clear answer received, but operational evaluations seem to dominate.  
Source: interviews.



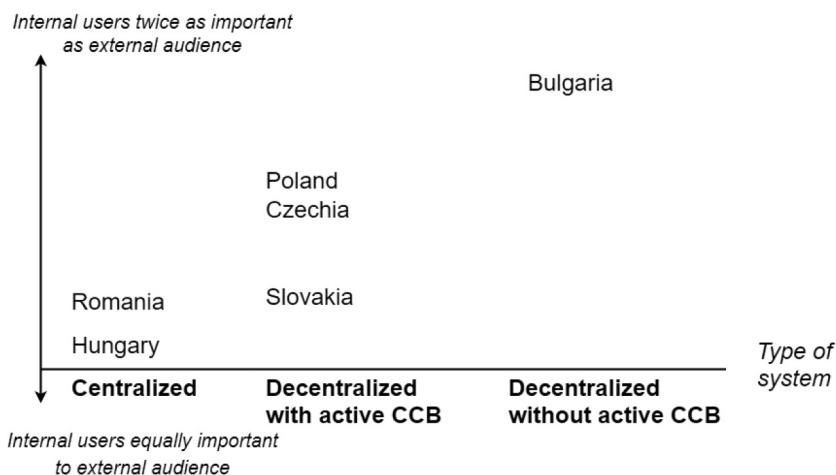
**Figure 7.** Share of impact-oriented studies in the total of completed evaluation studies (%).  
Source: own elaboration based on EU library of Member States CP evaluations.

(strategic focus in our framework) are in a minority in all countries. However, their share in the centralized systems of Romania and Hungary is significantly higher than that in any decentralized system.

Based on current data, it cannot be determined how exactly the different types of evaluations are used in studied countries. However, the EC pays attention almost solely to impact evaluations generated in Member States, which suggests that in centralized systems evaluation may serve accountability more often than in decentralized ones.

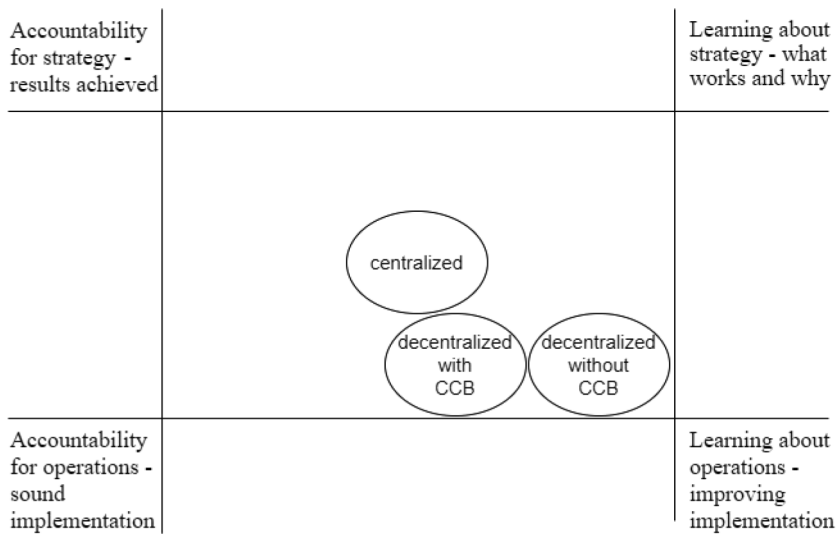
Our main measure of the evaluation system's dominant purpose is the perceived intended audience of evaluation units. In Hungary, external audiences – other institutions in the CP implementation system, domestic public institutions dealing with other policies, and institutions at the EU level – are as important as internal users – managers of other units and senior public administration staff in the same institution. This suggests that both evaluation purposes – accountability and learning – are equally pursued there. In all other countries, internal users are more important. The focus on internal users is most apparent in Poland's most decentralized evaluation system and the loose structure lacking an active CCB in Bulgaria. Therefore, decentralized systems seem to focus on learning purpose (Figure 8).

Summing up the findings on the three identified types of national CP evaluation system structure, there is a pattern in the observed similarities and differences in practice. Decentralized systems tend to produce more evaluation studies. Evaluation units in those systems are oriented on internal users, i.e. programme managers and senior public administration staff from the same organization. The large majority of acquired knowledge is focused on operations, it concerns the implementation process and procedures. This is true for both types of decentralized systems – with and without an active CCB – but the orientation on internal users is more apparent in the latter. Centralized systems produce fewer evaluation studies but provide a higher share of knowledge focused on strategy. Findings from the evaluations are intended for both internal and external users (e.g. institutions at the EU level). Referring these observations to our theoretical framework (see Figure 1), it can be stated that decentralized systems are oriented towards learning about operations and improving the implementation process. In centralized systems, learning about operations is also performed, but there is more attention on accountability for strategy – results achieved (Figure 9).



**Figure 8.** Target audience of evaluation findings.

Source: survey.



**Figure 9.** Relationship between the structure and functions of evaluation systems.  
*Source:* own elaboration.

Conclusions

This article has identified the variance in national CP evaluation systems’ organizational structures and has verified whether this variance correlates with differences in the systems’ functions. A comparative analysis has been conducted for six national CP evaluation systems of countries that joined the EU between 2004 and 2007.

The results show that there are visible differences in the structure of evaluation systems. Three basic types are identified: (1) centralized – one evaluation unit serving a multitude of user-organizations; (2) decentralized with an active CCB – several evaluation units spread among different ministries and one central unit performing a coordinating role; and (3) decentralized without an active CCB.

The study shows that the practice of evaluation units functioning in the systems of these three types differs in a manner indicating focus on different functions. Decentralized systems are oriented towards learning about operations only, while centralized systems also serve accountability for results.

On the practical side, these findings provide three insights for the structural arrangements of complex evaluation systems. First, any authority responsible for supervising or formulating a regulatory framework for evaluation systems (the EC in our case) should bear in mind that if certain evaluation functions are to be effectively fulfilled, the appropriate system structure must be indicated or required through regulations or guidelines.

Second, all analysed structures have their strengths and limitations. Decentralized systems respond better to user needs and support learning, but are mostly operational. There are limited strategic insights for accountability or learning about what works and why.

Evaluation units of centralized systems provide more independent impact evaluations suited to accountability, but their outcomes may not be relevant and useful for potential users within implementing organizations.

Third, if both accountability and learning are desired, the evaluation system needs some degree of decentralization on the one hand and the presence of an active CCB on the other. Decentralized evaluation units located within the same organization as their potential users are best able to provide relevant and timely operational knowledge. The CCB: (1) may secure accountability because of its independence; (2) has capacity to conduct robust impact studies; and (3) is in the best position to guide evaluation units and combine their findings into the reliable knowledge needed for strategic learning.

As regards implications for the theory, the study demonstrates a relationship between the structure and functions of the evaluation system but does not establish causality. This limitation opens the perspective for further studies. Based on current observations from the studied evaluation systems, it can be conjectured that it is the structure that influences functions and that other contextual factors lie behind decisions on structure. However, this should be verified empirically.

The issue of factors shaping structure is also open for further exploration. At this point, we may only speculate that these factors could include the formal and informal characteristics of public administration settings, cultures and traditions (Curristine et al., 2007), the nature of external pressure on national administrations (Højlund, 2014) or the power struggle between actors in the system (Martinaitis et al., 2018).

Furthermore, this study opens up the possibility for a more systematic analysis of diversity in systems performance in terms of independence, quality of findings and the specific structures' ability to combine individual studies into streams of evidence.

Finally, the current study was limited in geographical and policy terms (six CEECs, CP). Therefore, the article's typology should be further tested in comparative research on evaluation systems in other countries and specific policies. It could pave the way for a better understanding of the role of evaluation in public administration and policy practice.

## **Acknowledgements**

The authors thank the representatives of the studied countries' administrations, who facilitated the study's execution, and Weronika Orkisz Felcis, who contributed to the data collection process. Karol Olejniczak acknowledges the Fulbright Program – The Senior Award 2020–21 scholarship, which allowed completion of the final version of the article.

## **Declaration of conflicting interests**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## **Funding**


The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Empirical data was collected during a project commissioned by the

Ministry of Development Funds – Poland. The comparative analysis was conducted as part of the project financed by the National Science Centre (2019/33/B/HS5/01336).

## ORCID iDs

Tomasz Kupiec  <https://orcid.org/0000-0002-6469-7746>

Dominika Wojtowicz  <https://orcid.org/0000-0002-8433-8935>

Karol Olejniczak  <https://orcid.org/0000-0002-7079-2440>

## References

- Ágh A (2002) Public sector reforms, institutional design and strategy for good governance in east Central Europe. *Studies in East European Thought* 53(3): 233–255.
- Ágh A (2016) The decline of democracy in east-Central Europe. *Problems of Post-Communism* 63(5–6): 277–287.
- Argyris C and Schon DA (1995) *Organizational Learning II: Theory, Method, and Practice*. Massachusetts: FT Press.
- Balla S, Lodge M and Page E (2015) What makes a classic? Identifying and revisiting the classics of public policy and administration. In: Balla S, Lodge M and Page E (eds) *The Oxford Handbook of Classics in Public Policy and Administration*. Oxford: Oxford University Press, pp. 1–11.
- Batterbury SCE (2006) Principles and purposes of European Union Cohesion Policy evaluation. *Regional Studies* 40(2): 179–188.
- Chelimsky E and Shadish WR (1997) *Evaluation for the 21st Century: A Handbook*. London: Sage.
- Curristine T, Lonti Z and Joumard I (2007) Improving public sector efficiency: Challenges and opportunities. *OECD Journal on Budgeting* 7(1): 1–41.
- Dąbrowski M, Bachtler J and Bafoif F (2014) Challenges of multi-level governance and partnership: Drawing lessons from European Union cohesion policy. *European Urban and Regional Studies* 21(4): 355–363.
- Derlien HU (1990) Genesis and structure of evaluation efforts in comparative perspective. In Rist RC (ed.) *Program Evaluation and the Management of Government*. New Brunswick, NJ: Transaction, pp. 147–177.
- Dimitrov V, Goetz K and Wollmann H (2006) *Governing After Communism*. Boulder, CO: Rowman and Littlefield.
- Donaldson S, Patton M, Fetterman D, et al. (2010) The 2009 Claremont debates: The promise and pitfalls of utilization-focused and empowerment evaluation. *Journal of MultiDisciplinary Evaluation* 6(13): 15–57.
- Dubnick M and Frederickson G (2011). *Public Accountability: Performance Measurement, the Extended State, and the Search for Trust*. Washington, DC: National Academy of Public Administration.
- Dyas GP and Thanheiser H (1976) *The Emerging European Enterprise*. London: Macmillan.
- Favoreu C, Carassus D, Gardey D, et al. (2015) Performance management in the local public sector in France: An administrative rather than a political model. *International Review of Administrative Sciences* 81(4): 672–693.
- Fournier DM (2005) Evaluation. In: Mathison S (ed.) *Encyclopedia of Evaluation*. Thousand Oaks, CA: Sage, pp. 139–140.
- Fratesi U and Wislade F (2017) The impact of European Cohesion Policy in different contexts. *Regional Studies* 51(6): 817–821.

- Furubo JE, Rist RC and Sandahl R (2002) *International Atlas of Evaluation*. New Brunswick, NJ: Transaction.
- Hall DJ and Saias MA (1980) Strategy follows structure. *Strategic Management Journal* 1(2): 149–163.
- Hanberger A (2011) The real functions of evaluation and response systems. *Evaluation* 17(4): 327–349.
- Harris IC and Ruefli TW (2000) The strategy/structure debate: An examination of the performance implications. *Journal of Management Studies* 37(2): 587–604.
- Henry GT and Mark MM (2003) Beyond Use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation* 24(3): 293–314.
- Højlund S (2014) Evaluation use in the organisational context – Changing focus to improve theory. *Evaluation* 20(1): 26–43.
- Kudo H (2003) Between the 'governance' model and the Policy Evaluation Act: New public management in Japan. *International Review of Administrative Sciences* 69(4): 483–504.
- Kupiec T (2020) Do we learn how to learn? Evolution of functions of evaluation systems – The case of Polish regional administration. *Studia Regionalne i Lokalne* 81(3): 7–22.
- Kusek JZ and Rist RC (2004) *Ten Steps to a Result-based Monitoring and Evaluation System*. Washington, DC: World Bank.
- Leeuw FL and Furubo JE (2008) Evaluation systems: What are they and why study them? *Evaluation* 14(2): 157–169.
- Lijphart A (1971) Comparative politics and the comparative method. *American Political Science Review* 65(3): 682–693.
- Liverani A and Lundgren HE (2007) Evaluation systems in development aid agencies: An analysis of DAC peer reviews 1996–2004. *Evaluation* 13(2): 241–256.
- Mark MM, Greene J and Shaw I (2006) The evaluation of policies, programs, and practices. In: Shaw I, Greene J and Mark M (eds) *The Sage Handbook of Evaluation*. London: Sage, pp. 1–30.
- Mark MM, Henry GT and Julnes G (1999) Toward an integrative framework for evaluation practice. *American Journal of Evaluation* 20(2): 177–198.
- Martinaitis Ž, Christenko A and Kraučūnienė L (2018) Evaluation systems: How do they frame, generate and use evidence? *Evaluation* 25(1): 46–61.
- Meyer-Sahling JH (2009) Varieties of legacies: A critical review of legacy explanations of public administration reform in east central Europe. *International Review of Administrative Sciences* 75(3): 509–527.
- Mintzberg H (1992) *Structure in Fives: Designing Effective Organizations*. Upper Saddle River, NJ: Prentice Hall.
- OECD (1998) Best practice guidelines for evaluation (PUMA policy brief No. 5). Paris: Organisation for Economic Co-operation and Development. Available at: <http://www.oecd.org/governance/budgeting/1902965.pdf> (accessed 13 August 2020).
- Oh Y and Lee K (2020) External control mechanisms and red tape: Testing the roles of external audit and evaluation on red tape in quasi-governmental organizations. *International Review of Administrative Sciences*.
- Olejniczak K (2013) Mechanisms shaping evaluation system – A case study of Poland 1999–2010. *Europe-Asia Studies* 65(8): 1642–1666.
- Olejniczak K, Raimondo E and Kupiec T (2016) Evaluation units as knowledge brokers: Testing and calibrating an innovative framework. *Evaluation* 22(1): 168–189.
- Olejniczak K, Kupiec T and Newcomer K (2017) Learning from evaluation – the knowledge users; perspective. *Evaluation Theory and Practice* 5(2): 49–74.
- Ostrom E (2007) *Governing the Commons: The Evolution of Institutions for Collective Action*. New York: Cambridge University Press.

- Patton MQ (2004) Evaluation research. In: Lewis-Beck MS, Bryman A and Liao TF (eds) *The Sage Encyclopedia of Social Science Research Methods*. Thousand Oaks, CA: Sage, pp. 337–340.
- Picciotto R (2016) Evaluation and bureaucracy: The tricky rectangle. *Evaluation* 22(4): 424–434.
- Raimondo E (2018) The power and dysfunctions of evaluation systems in international organizations. *Evaluation* 24(1): 26–41.
- Rossi P, Lipsey MW and Freeman H (2004) *Evaluation: A Systematic Approach*. London: Sage.
- Sartori G (1991) Comparing and miscomparing. *Journal of Theoretical Politics* 3(3): 243–257.
- Speer S, Pattyn V and De Peuter B (2015) The growing role of evaluation in parliaments: Holding governments accountable? *International Review of Administrative Sciences* 81(1): 37–57.
- Stockmann R, Meyer W and Taube L (eds) (2020) *The Institutionalisation of Evaluation in Europe*. Cham: Palgrave Macmillan.
- Thijs N, Hammerschmid G and Palaric E (2017) *A Comparative Overview of Public Administration Characteristics and Performance in EU28*. Brussels: European Commission.
- Van Der Meer FB and Edelenbos J (2006) Evaluation in multi-actor policy processes: Accountability, learning and Co-operation. *Evaluation* 12(2): 201–218.
- Widmer T and Neuenschwander P (2004) Embedding evaluation in the Swiss federal administration. Purpose, institutional design and utilization. *Evaluation* 10(4): 388–409.
- Williams B and Imam I (2007) *Systems Concepts in Evaluation: An Expert Anthology*. Point Reyes: EdgePress of Inverness.
- Zwaan P, Van Voorst S and Mastenbroek E (2016) Ex post legislative evaluation in the European Union: Questioning the usage of evaluations as instruments for accountability. *International Review of Administrative Sciences* 82(4): 674–693.

**Tomasz Kupiec** is an assistant professor at The Centre for European Regional and Local Studies (EUROREG), University of Warsaw, Poland. His main research interest are performance of evaluation systems and use of knowledge in public administration.

**Dominika Wojtowicz** is a professor and researcher at the Department of Economics of Kozminski University. Her interests focus on the effectiveness of public interventions aimed at regional and local development.

**Karol Olejniczak** is a professor of public policy at SWPS University, Poland, and a co-founder of research company Evaluation for Government Organizations. His work focuses on programme evaluation and behavioural policy design.